HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P. O. Box 272400
Fort Collins, Colorado 80527-2400

PATENT APPLICATION

ATTORNEY DOCKET NO. 200301917-1

IN THE
UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor(s): Marc A. NAJORK

Confirmation No.: 6430

Application No.: 09/706198

Examiner: Brian Goddhard

Filing Date: Nov 03, 2000

Group Art Unit: 2161

Title: SYSTEM AND METHOD FOR DISTRIBUTED WEB CRAWLING

Mail Stop Appeal Brief-Patents
Commissioner For Patents
PO Box 1450
Alexandria, VA 22313-1450

TRANSMITTAL OF APPEAL BRIEF

Sir:

Transmitted herewith is the Appeal Brief in this application with respect to the Notice of Appeal filed on April 4, 2005.

The fee for filing this Appeal Brief is (37 CFR 1.17(c)) $500.00.

(complete (a) or (b) as applicable)

The proceedings herein are for a patent application and the provisions of 37 CFR 1.136(a) apply.

( ) (a) Applicant petitions for an extension of time under 37 CFR 1.136 (fees: 37 CFR 1.17(a)-(d) for the total number of months checked below:

( )	one month	$120.00
( )	two months	$450.00
( )	three months	$1020.00
( )	four months	$1590.00

06/07/2005 HTECKLU1 00000056 082025 09706198
01 FC:1402 500.00 DA

( ) The extension fee has already been filled in this application.

(X) (b) Applicant believes that no extension of time is required. However, this conditional petition is being made to provide for the possibility that applicant has inadvertently overlooked the need for a petition and fee for extension of time.

Please charge to Deposit Account **08-2025** the sum of **$500.00**. At any time during the pendency of this application, please charge any fees required or credit any over payment to Deposit Account 08-2025 pursuant to 37 CFR 1.25. Additionally please charge any fees to Deposit Account 08-2025 under 37 CFR 1.16 through 1.21 inclusive, and any other sections in Title 37 of the Code of Federal Regulations that may regulate fees. A duplicate copy of this sheet is enclosed.

( ) I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Alexandria, VA 22313-1450. Date of Deposit:_____

OR

(X) I hereby certify that this paper is being transmitted to the Patent and Trademark Office facsimile number (703) 872-9306 on 6/6/2005

Number of pages: 25

Typed Name: Be Henry

Signature: Be Henry

Respectfully submitted,

Marc A. NAJORK

By _____

Philip Lyren

Attorney/Agent for Applicant(s)
Reg. No. 40,709

Date: 6/6/2005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Marc A. Najork	Examiner: Brian D. Goddhard
Serial No.:	09/706,198	Group Art Unit: 2161
Filed:	November 3, 2000	Docket No.: 200301917-1
Title:	System and Method for Distributed Web Crawling	

APPEAL BRIEF UNDER 37 C.F.R. § 41.37

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

This Appeal Brief is filed in response to the Final Office Action mailed January 14, 2005 and the Notice of Appeal filed on April 4, 2005.

AUTHORIZATION TO DEBIT ACCOUNT

It is believed that no extensions of time or fees are required, beyond those that may otherwise be provided for in documents accompanying this paper. However, in the event that additional extensions of time are necessary to allow consideration of this paper, such extensions are hereby petitioned under 37 C.F.R. § 1.136(a), and any fees required (including fees for net addition of claims) are hereby authorized to be charged to Hewlett-Packard Development Company's deposit account no. 08-2025.

I. REAL PARTY IN INTEREST

The real party-in-interest is the assignee, Hewlett-Packard Company, a Delaware corporation, having its principal place of business in Palo Alto, California.

II. RELATED APPEALS AND INTERFERENCES

There are no known related appeals or interferences known to appellant, the appellant's legal representative, or assignee that will directly affect or be directly affected by or have a bearing on the Appeal Board's decision in the pending appeal.

III. STATUS OF CLAIMS

Claims 1 - 4 and 6 – 18 stand finally rejected. No claims have been allowed. The final rejection of claims 1 - 4 and 6 – 18 is appealed.

IV. STATUS OF AMENDMENTS

In response to the Final Office Action, no claims were amended. The claims on appeal and in the following Claim Appendix VIII correspond to the presently pending claims.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The summary is set forth in three exemplary embodiments that correspond to independent claims 1, 6, and 10. Discussions about elements and recitations of these claims can be found at least at the cited locations in the specification and drawings.

Claim 1

A method of downloading data sets by a plurality of web crawlers from among a plurality of host computers, comprising the steps of:

assigning a web crawler identifier to each one of the plurality of web crawlers; (FIG. 1: page 4, lines 22-27; FIG. 4: page 7, line 1 – page 9, line 11)

for each respective web crawler: (see FIG. 3)

downloading at least one data set that includes addresses of one of more referred data sets; (FIG. 3: page 6, lines 1-20)

identifying the addresses of the one or more referred data sets, wherein each identified address includes a host computer identifier; (FIG. 3: page 6, lines 1-31; FIG. 4: page 7, lines 1-22)

for each identified address: (see FIGS. 3 and 4)

generating a representation of the host computer identifier; (FIG. 4: page 7, line 11 – page 8, line 15)

determining a web crawler identifier to which the representation corresponds; and (FIG. 4: page 8, lines 17-33)

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned. (FIG. 4, page 8, line 17 – page 9, line 11; Examples at blocks 164-166 in FIG. 4)

Claim 6

A web crawler system for downloading data set addresses from among a plurality of host computers, comprising:

a plurality of web crawlers, wherein each web crawler has been assigned a web crawler identifier; (FIG. 1: page 4, lines 22-27)

for each respective web crawler: (see FIG. 3)

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and (FIG. 1: page 3, line 30 – page 4, line 27; FIG. 2: page 4, line 28 – page 5, lines 32; FIG. 3: page 6, lines 1-31; FIG. 4: page 7, lines 1-22)

an address distribution module for processing the identified addresses, the address distribution module including instructions for: (see FIG. 4)

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers; (FIG. 4: page 7, line 11 – page 8, line 15)

determining a web crawler identifier to which the representation corresponds; and (FIG. 4: page 8, lines 17-33)

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned. (FIG. 4: page 8, line 17 – page 9, line 11; Examples at blocks 164-166 in FIG. 4)

Claim 10

A computer program product for use in conjunction with a web crawler system wherein each web crawler is assigned a web crawler identifier, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising: (Page 13, lines 7-14; FIG. 1: page 4, lines 22-27; FIG. 4: page 7, line 1 – page 9, line 11)

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and (FIG. 1: page 3, line 30 – page 4, line 27; FIG. 2: page 4, line 28 -- page 5, lines 32; FIG. 3: page 6, lines 1-31; FIG. 4: page 7, lines 1-22)

an address distribution module for processing the identified addresses, the address distribution module including instructions for: (see FIG. 4)

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers; (FIG. 4: page 7, line 11 – page 8, line 15)

determining a web crawler identifier to which the representation corresponds; and (FIG. 4: page 8, lines 17-33)

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned. (FIG. 4: page 8, line 17 – page 9, line 11; Examples at blocks 164-166 in FIG. 4)

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

(1) Claims 1, 6, and 10 are rejected under 35 USC § 102(a) as being anticipated by an article entitled "Mercator: A scalable, extensive Web Crawler" by Heydon et al. (hereinafter Heydon).

(2) Claims 1-4 and 6-14 are rejected under 35 USC § 102(e) as being anticipated by Najork et al. (USPN 6,377,984, hereinafter Najork).

(3) Claims 1-4 and 6-14 are rejected under 35 USC § 102(f) because applicant did not invent the claimed subject matter.

(4) Claims 1-4 and 6-14 are rejected under 35 USC § 102(e) as being anticipated by Eichstaedt et al. (USPN 6,182,085, hereinafter Eichstaedt).

(5) Claims 15-18 are rejected under 35 USC § 103 as being unpatentable over Eichstaedt in view of Najork.

VII. ARGUMENT

(1) Claim Rejections: 35 USC § 102(a)

Claims 1, 6, and 10 are rejected under 35 USC § 102(a) as being anticipated by an article entitled "Mercator: A scalable, extensive Web Crawler" by Heydon et al. (hereinafter Heydon). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102 requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Heydon neither teaches nor suggests each element in claims 1, 6, and 10, these claims are allowable over Heydon.

Response to Final Office Action

Applicant has repeatedly argued that Heydon does not teach or suggest a "plurality of web crawlers." The Final Office Action disagrees and states:

> Heydon's worker threads are each considered separate "web crawlers" as each thread performs the function of a "web crawler." Nowhere has applicant provided a definition of "web crawler" that goes above and beyond the conventional definition, that would distinguish from Heydon's worker thread. (Final OA, pages 13-14).

Applicant respectfully disagrees. Applicant uses the terms "web crawler" and "thread" in the plain meaning given to these terms per one of ordinary skill in the art (see MPEP 2111.01: Words of claim must be given their plain meaning). Webopedia (see www.webopedia.com) is an online dictionary dedicated to defining computer and internet related terms. Webopedia defines "web crawler" and "thread" as follows:

> **Web crawler** (note: web crawler and spider are synonyms):
>
> A program that automatically fetches Web pages. Spiders are used to feed pages to search engines. It's called a spider because it *crawls* over the Web. Another term for these programs is *webcrawler*.
>
> **Thread:**
>
> (2) In programming, a part of a program that can execute independently of other parts. Operating systems that support multithreading enable programmers to design programs whose threaded parts can execute concurrently.

Thus, a "web crawler" **is a program** that automatically fetches web pages. A "thread" is **part of a program** that can execute independently of other parts. Even Applicant's specification states: "A web crawler is a program that automatically finds and downloads documents from host computers in an Intranet or the world wide web" (see p. 1, lines 24-25).

Applicant respectfully asserts that the Office Action has not applied the terms "web crawler" and "thread" in accordance with their plain meaning. As noted above, the Office Action equates Heydon's "threads" as being "web crawlers" (see quote above of Final OA, pages 13-14: "Heydon's worker threads are each considered separate web crawlers...."). The Office Action utilizes the terms "web crawler" and "thread" in a manner that is repugnant to the plain meaning of these terms.

Claim 1

Independent claim 1 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 1 recites "a plurality of web crawlers." By contrast, Heydon does not teach or suggest a plurality of web crawlers. Heydon teaches a **single** web crawler (see Abstract: "This paper describes Mercator, a scalable, extensible web crawler ...."). Section 3.1 paragraph 1 does state: "Crawling is performed by multiple

worker threads." Multiple worker threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Heydon teaches a **single** web crawler.

As another example, claim 1 recites "assigning a web crawler identifier to each one of the plurality of web crawlers." Heydon does not teach or suggest this limitation. The Office Action cites Section 3.2, third paragraph of Heydon for teaching this limitation. This section of Heydon teaches that Mercators' URL frontier includes distinct FIFO subqueues; one FIFO subqueue per worker thread. This section further states: "Second, when a new URL is added, the FIFO subqueue in which it is placed is determined by the URL's canonical host name." Nowhere does this section teach or suggest assigning an identifier to each one of a plurality of web crawlers.

As another example, claim 1 recites:

> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 1, though, are not shown or suggested.

Claim 6

Independent claim 6 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 6 recites "a plurality of web crawlers." By contrast, Heydon does not teach or suggest a plurality of web crawlers. Heydon teaches a **single** web crawler (see Abstract: "This paper describes Mercator, a scalable, extensible web crawler ...."). Section 3.1 paragraph 1 does state: "Crawling is performed by multiple worker threads." Multiple worker threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Heydon teaches a **single** web crawler.

As another example, claim 6 recites "wherein each web crawler has been assigned a web crawler identifier." Heydon does not teach or suggest this limitation. The Office Action cites Section 3.2, third paragraph of Heydon for teaching this limitation. This section of Heydon teaches that Mercators' URL frontier includes distinct FIFO subqueues; one FIFO subqueue per worker thread. This section further states: "Second, when a new URL is added, the FIFO subqueue in which it is placed is determined by the URL's canonical host name." Nowhere does this section teach or suggest a plurality of web crawlers with each web crawler assigned a web crawler identifier.

As another example, claim 6 recites:

> for each respective web crawler:
>
> a main web crawler module ...
>
> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 6, though, are not shown or suggested.

Claim 10

Independent claim 10 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 10 recites:

> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising

> the web crawler to which the determined web crawler identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 10, though, are not shown or suggested in Heydon.

(2) Claim Rejections: 35 USC § 102(e)

Claims 1-4 and 6-14 are rejected under 35 USC § 102(e) as being anticipated by Najork et al. (USPN 6,377,984, hereinafter Najork). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102(e) requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Najork neither teaches nor suggests each element in claims 1-4 and 6-14, these claims are allowable over Najork.

Response to Final Office Action

In Applicant's response dated 15 July 2004, Applicant argues that Njork does not teach or suggest a "plurality of web crawlers." The Final Office Action disagrees and states:

> The Examiner disagrees for the same reasons discussed above with regard to Heydon. Namely, each of Najork's worker threads is considered equivalent to a "web crawler" as claimed. (Final OA, page 14).

Applicant restates the argument above in Argument Section (1): The Office Action has not applied the terms "web crawler" and "thread" in accordance with their plain meaning. The Office Action equates Njork's "threads" as being "web crawlers" (see quote above of Final OA, page 14: "Namely, each of Najork's worker threads is considered equivalent to a "web crawler" as claimed."). The Office Action utilizes the

terms "web crawler" and "thread" in a manner that is repugnant to the plain meaning of these terms. Please see the dictionary definition stated above for the terms "web crawler" and "thread" in Argument Section (1).

Claim 1

Independent claim 1 recites numerous limitations that are not taught or suggested in Najork. For example, claim 1 recites "a plurality of web crawlers." By contrast, Najork does not teach or suggest a plurality of web crawlers. Najork teaches a **single** web crawler:

> FIG. 1 shows an exemplary embodiment of a distributed computer system 100. The distributed computer system 100 includes **a web crawler 102** connected to a network 103 through a network interconnection 110. (Col. 3, lines 60-63: emphasis added).

Fig. 1 of Najork shows a single web crawler 102 with memory 118 that includes threads 130. Multiple threads, though, are **not** a plurality of web crawlers as recited in claim 1. In fact, Fig. 1 of Najork and the accompanying description clearly teach and suggest a **single** web crawler.

As another example, claim 1 recites "assigning a web crawler identifier to each one of the plurality of web crawlers." Najork does not teach or suggest this limitation. The Office Action cites identifier "r" (Figs. 2-4) to teach one of a plurality of web crawlers (each thread being a crawler, see also Fig. 3B). These figures and accompanying description do not teach or suggest assigning an identifier to each one of a plurality of web crawlers. By contrast, the figures are generally directed to FIFO queues for a single web crawler.

As yet another example, claim 1 recites:

> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the

> identified address to the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO queues for URLs downloaded within a single web crawler. The claimed limitations in claim 1, though, are not shown or suggested.

Dependent claims 2-4 depend from claim 1 and thus inherit all the limitations of base claim 1. Thus, for at least the reasons given in connection with independent claim 1, dependent claims 2-4 are also allowable over Najork.

Claim 6

Independent claim 6 recites numerous limitations that are not taught or suggested in Najork. For example, claim 6 recites "a plurality of web crawlers." By contrast, Najork does not teach or suggest a plurality of web crawlers. Najork teaches a **single** web crawler:

> FIG. 1 shows an exemplary embodiment of a distributed computer system 100. The distributed computer system 100 includes **a web crawler 102** connected to a network 103 through a network interconnection 110. (Col. 3, lines 60-63: emphasis added).

Fig. 1 of Najork shows a single web crawler 102 with memory 118 that includes threads 130. Multiple threads, though, are **not** a plurality of web crawlers as recited in claim 1. In fact, Fig. 1 of Najork and the accompanying description clearly teach and suggest a **single** web crawler.

As another example, claim 6 recites "wherein each web crawler has been assigned a web crawler identifier." Najork does not teach or suggest this limitation. The Office Action cites identifier "r" (Figs. 2-4) to each one of a plurality of web crawlers (each thread being a crawler, see also Fig. 3B). These figures and accompanying description do not teach or suggest assigning an identifier to each one of a plurality of web crawlers. By contrast, the figures are generally directed to FIFO queues for a single web crawler.

Claim 6 recites numerous other recitations that are not taught or suggested in Najork. For example, claim 6 recites:

> for each respective web crawler:
>
> a main web crawler module ...
>
> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO queues for URLs downloaded within a single web crawler. The claimed limitations in claim 6, though, are not shown or suggested.

Dependent claims 7-9 depend from claim 6 and thus inherit all the limitations of base claim 6. Thus, for at least the reasons given in connection with independent claim 6, dependent claims 7-9 are also allowable over Najork.

Claim 10

Independent claim 10 recites numerous limitations that are not taught or suggested in Najork. For example, claim 10 recites:

> determining a web crawler identifier to which the representation corresponds; and
>
> when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO

queues for URLs downloaded within a single web crawler. The claimed limitations in claim 10 are not shown or suggested in Najork.

Dependent claims 11-14 depend from claim 10 and thus inherit all the limitations of base claim 10. Thus, for at least the reasons given in connection with independent claim 10, dependent claims 11-14 are also allowable over Najork.

(3) Claim Rejections: 35 USC § 102(f)

Claims 1-4 and 6-14 are rejected under 35 USC § 102(f) because applicant did not invent the claimed subject matter. This rejection is traversed.

First, a Declaration for Patent Application was concurrently submitted with the filing of the patent application on November 3, 2000. This declaration was entered by the U.S. Patent Office and forms part of the file history for this patent application. The Examiner has not raised any objections to the declaration. In the declaration, the inventor (Marc A. Najork) states that he is the original, first, and sole inventor of the claimed subject matter.

Second, the Office Action argues the following (see FOA at page):

> The claimed invention is fully disclosed in the article entitled "Mercator: A scalable, extensible Web crawler" by Heydon et al. and U.S. Patent No. 6,377,984 to Najork et al. as shown above. While applicant appears as party to both references (co-author of the article and co-inventor of the '984 Patent), at least one other author/inventor are party to each reference as well, showing that applicant did not invent the claimed subject matter alone.

Applicant respectfully disagrees with the application of law and conclusions of the Examiner. In Argument Sections (1) and (2) above, Applicant demonstrates that the claimed subject matter is patentable over the teachings and suggestions in Heydon and/or Najork. In other words, Heydon and/or Najork do not teach or suggest all of the elements of the pending claims. Applicant alone invented the claimed subject matter that is not taught or suggested in Heydon and/or Najork.

(4) Claim Rejections: 35 USC § 102(e)

Claims 1-4 and 6-14 are rejected under 35 USC § 102(e) as being anticipated by Eichstaedt et al. (USPN 6,182,085, hereinafter Eichstaedt). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102(e) requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Eichstaedt neither teaches nor suggests each element in claims 1-4 and 6-14, these claims are allowable over Eichstaedt.

Claim 1

Independent claim 1 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 1 recites "assigning a web crawler identifier to each one of the plurality of web crawlers." Eichstaedt does not teach or suggest this limitation. The Office Action cites Col. 10 and gatherer processor id "i" as teaching this limitation. This section of Eichstaedt teaches how to partition the web-graph among numerous gatherers or processors:

> Assuming that one version of the present invention has k "gatherers" or processors, the web-graph is divided into k sub-graphs $W_1, \ldots W_k$. Each sub-graph is mapped to a processor (e.g., W_i to processor i). (Col. 10, lines 18-21).

Thus, this section of Eichstaedt teaches how to divide the web-graph between processors. This section does not teach or suggest assigning a web crawler identifier to each processor or gatherer.

As another example, claim 1 recites downloading data sets and identifying addresses of one or more referred data sets. For each identified address, claim 1 specifically recites "generating a representation of the host computer identifier" and "determining a web crawler identifier to which the representation corresponds." Eichstaedt does not teach these limitations. As noted, Eichstaedt does not assign web crawler identifiers to each gatherer or processor. As such, Eichstaedt does not generate a

representation of a host computer identifier and then determine a web crawler identifier to which the representation corresponds.

The Office Action relies on Fig. 6 and Col. 6 (for example, lines 39-67 and 2-38). These sections in Eichstaedt teach dividing the web-space (the URL space) into sub-spaces and assigning sub-spaces to certain processors (Col. 6, lines 30-32). When new URLs are added to the web-space, the processor processes URLs belonging to its sub-space and routes other URLs (i.e., those not belonging to its sub-space) to the proper processor (Col. 6, lines 33-38). Notice though, that Eichstaedt does not generate a representation of a host computer identifier and then determine a web crawler identifier to which the representation corresponds.

Dependent claims 2-4 depend from claim 1 and thus inherit all the limitations of base claim 1. Thus, for at least the reasons given in connection with independent claim 1, dependent claims 2-4 are also allowable over Eichstaedt.

Claim 6

Independent claim 6 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 6 recites "wherein each web crawler has been assigned a web crawler identifier." For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

As another example, claim 6 recites a main web crawler module that identifies addresses of referred data sets, wherein each identified address includes a host computer identifier. An address distribution module processes the identified addresses and includes instructions for "generating a representation of the host computer identifier" and "determining a web crawler identifier to which the representation corresponds." For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

Dependent claims 7-9 depend from claim 6 and thus inherit all the limitations of base claim 6. Thus, for at least the reasons given in connection with independent claim 6, dependent claims 7-9 are also allowable over Eichstaedt.

Claim 10

Independent claim 10 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 10 recites "wherein each web crawler has been assigned a web crawler identifier." For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

As another example, claim 10 recites a main web crawler module that identifies addresses of referred data sets, wherein each identified address includes a host computer identifier. An address distribution module processes the identified addresses and includes instructions for "generating a representation of the host computer identifier" and "determining a web crawler identifier to which the representation corresponds." For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

Dependent claims 11-14 depend from claim 10 and thus inherit all the limitations of base claim 10. Thus, for at least the reasons given in connection with independent claim 10, dependent claims 11-14 are also allowable over Eichstaedt.

(5) Claim Rejections: 35 USC § 103

Claims 15-18 are rejected under 35 USC § 103 as being unpatentable over Eichstaedt in view of Najork. Applicant respectfully traverses.

To establish a prima facie case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art cited must teach or suggest all the claim limitations. *See* M.P.E.P. § 2143. Applicant asserts that the rejection does not satisfy these criteria.

Claims 15-18 depend from independent claims 1, 6, and 10. For each of these independent claims, Argument Section (2) above discusses numerous claim elements that are not taught or suggested in Najork, and Argument Section (4) above discusses numerous claim elements that are not taught or suggested in Eichstaedt. The combination of Najork and Eichstaedt does not cure the noted deficiencies. In other words, for at least

the reasons noted in Argument Sections (2) and (4) above with respect to the independent claims, the combination of Najork and Eichstaedt does not teach or suggest all the claim limitations of dependent claims 15-18.

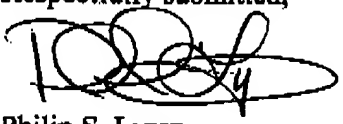
CONCLUSION

In view of the above, Appellant respectfully requests the Board of Appeals to reverse the Examiner's rejection of all pending claims.

Any inquiry regarding this Amendment and Response should be directed to Philip S. Lyren at Telephone No. (281) 514-8236, Facsimile No. (281) 514-8332. In addition, all correspondence should continue to be directed to the following address:

Hewlett-Packard Company
Intellectual Property Administration
P.O. Box 272400
Fort Collins, Colorado 80527-2400

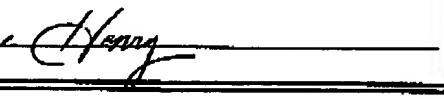
Respectfully submitted,



Philip S. Lyren
Reg. No. 40,709
Ph: 281-514-8236

CERTIFICATE UNDER 37 C.F.R. 1.8

The undersigned hereby certifies that this paper or papers, as described herein, is being transmitted to the United States Patent and Trademark Office facsimile number 703-872-9306 on this 6th day of June, 2005.

By 
Name: Be Henry

VIII. Claims Appendix

1. (original) A method of downloading data sets by a plurality of web crawlers from among a plurality of host computers, comprising the steps of:

assigning a web crawler identifier to each one of the plurality of web crawlers;

for each respective web crawler:

downloading at least one data set that includes addresses of one of more referred data sets;

identifying the addresses of the one or more referred data sets, wherein each identified address includes a host computer identifier;

for each identified address:

generating a representation of the host computer identifier;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned.

2. (original) The method of claim 1, wherein

the plurality of web crawlers consists of n web crawlers; and

generating the representation includes computing a hash function of the host computer identifier to generate an integer value that is a member of a set of n predefined distinct values.

3. (original) The method of claim 1, wherein

the plurality of web crawlers consists of n web crawlers; and

generating the representation includes computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

4. (original) The method of claim 1, wherein the sending step includes:

determining a web crawler address for the web crawler to which the determined web crawler identifier is assigned;

transmitting the identified data set address to the destination web crawler at the determined web crawler address.

5. (canceled)

6. (original) A web crawler system for downloading data set addresses from among a plurality of host computers, comprising:

a plurality of web crawlers, wherein each web crawler has been assigned a web crawler identifier;

for each respective web crawler:

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and

an address distribution module for processing the identified addresses, the address distribution module including instructions for:

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

7. (original) The web crawler system of claim 6 wherein

the plurality of web crawlers consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

8. (original) The web crawler system of claim 6, further comprising:

for each respective web crawler, a web crawler interface for transmitting the identified address to the destination web crawler and for receiving identified addresses from each of the plurality of web crawlers other than the respective web crawler.

9. (original) The web crawler system of claim 6, further comprising:

for each respective web crawler, a lookup table storing for each of the plurality of web crawler identifiers a corresponding web crawler address, said lookup table for use by the address distribution module in determining a web crawler address to which to send the identified data set address.

10. (original) A computer program product for use in conjunction with a web crawler system wherein each web crawler is assigned a web crawler identifier, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising:

a main web crawler module for downloading and processing data sets stored on a plurality of host computers, the main web crawler module identifying addresses of the one or more referred data sets in the downloaded data sets, wherein each identified address includes a host computer identifier; and

an address distribution module for processing the identified addresses, the address distribution module including instructions for:

generating a representation of the host computer identifier, wherein the representation corresponds to one of the web crawler identifiers;

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

11. (original) The computer program product of claim 10, wherein:

the web crawler system consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a function of the host computer identifier to generate an integer value that is a member of a set of n predefined distinct values.

12. (original) The computer program product of claim 10, wherein:

the web crawler system consists of n web crawlers; and

the address distribution module's instructions for generating the representation includes instructions for computing a hash function of the host computer identifier to generate an intermediate value V, and computing V modulo n.

13. (original) The computer program product of claim 10, further comprising:

a web crawler interface for transmitting the identified address to the destination web crawler and for receiving identified addresses from each of the plurality of web crawlers other than the respective web crawler.

14. (original) The computer program product of claim 10, further comprising:

a lookup table storing for each of the plurality of web crawler identifiers a corresponding web crawler address, said lookup table for use by the address distribution module in determining a web crawler address to which to send the identified data set address.

15. (previously presented) The method of claim 1, wherein each respective web crawler includes multiple threads to download and process documents from a plurality of host computers.

16. (previously presented) The web crawler system of claim 6 wherein each of the plurality of web crawlers includes multiple threads to download and process documents from a plurality of host computers.

17. (previously presented) The computer program product of claim 10 wherein each web crawler includes multiple threads.

18. (previously presented) The computer program product of claim 17 wherein each thread executes a main web crawler module.

IX. EVIDENCE APPENDIX

None.

X. RELATED PROCEEDINGS APPENDIX

None.